Decadal Changes in Subseasonal Predictability, Forecast Bias, and Model Calibration

Dan C. Collins[1]

1 NOAA Climate Prediction Center (CPC)
2 Innovim, LLC

Decadal Changes in Subseasonal Predictability, Forecast Bias, and Model Calibration

Retrospective forecasts for week-2, 2-meter temperature from the Global Ensemble System for 1985-2010 are used to estimate the model climatology, including the mean and variance, as well as the correlation with observations. The predictable signal is estimated from the variance of the ensemble mean and ensemble spread is used to estimate the uncertainty. Predictability is determined from both the true correlation to observations and the model-predicted correlations calculated from the signal and ensemble spread. Trends were fit to the predictability, both model-estimated and realized, as well as to the difference between the model climatology and the observed climatology, i.e. bias. The bias is shown to have clear trends that vary globally indicating that systematic bias between the ensemble mean and observations is changing as both the mean climate state and the observational analysis change. For purposes of model calibration of ensemble forecasts, bias correction must change with time. These results for week-2 are likely to be applicable for other subseasonal to seasonal prediction systems.

In operations, mean systematic bias-corrections and calibration from the full 1985-2010 retrospective GEFS dataset are applied, however changes in bias of real-time forecasts following 2010 are demonstrated. Verification for recent analyses to correct the residual bias using forecasts for 2011-2014 will be shown.

NOAA-ESRL retrospective forecasts

- Current operational NCEP Global Ensemble Forecast System (GEFS) as of February 2012

- T254L42 (about ½ degree grid spacing) in week 1 and T190L42 (about ¾-degree) in week 2

- Daily 0Z cycle 11 ensemble members,
    - 10 perturbations + control
    - 4 cycles x 21 members per day in real-time GEFS
- 1985-2010

- Initial conditions from Climate Forecast System Reanalysis (CFSR) (2011 and real-time using GDAS)

Week-2 reliability of reforecast-calibrated GEFS probabilities compared to NAEFS, bias-corrected GEFS and uncorrected GEFS
(3 years 2011-2013)
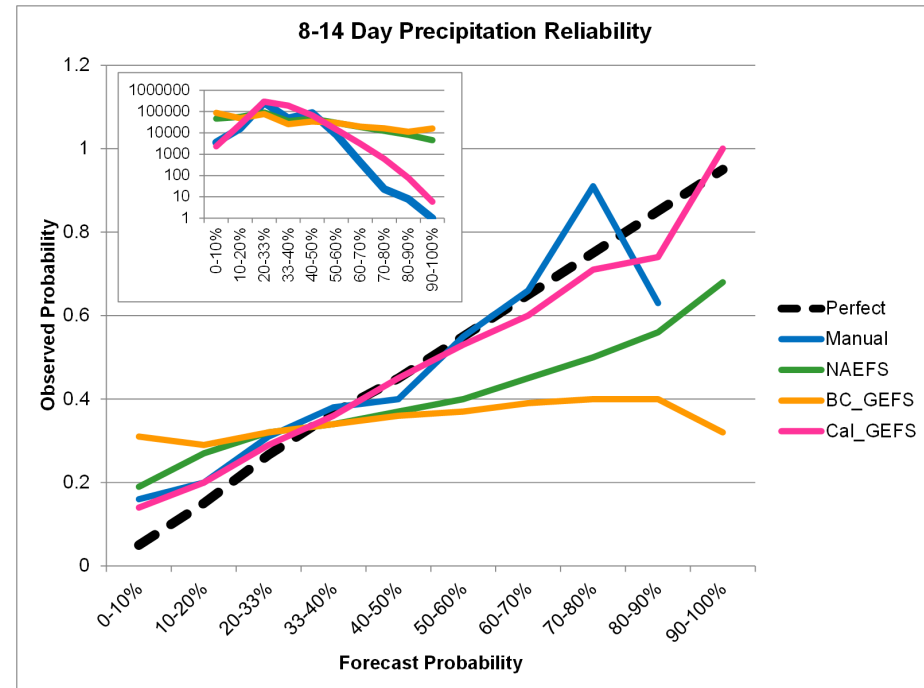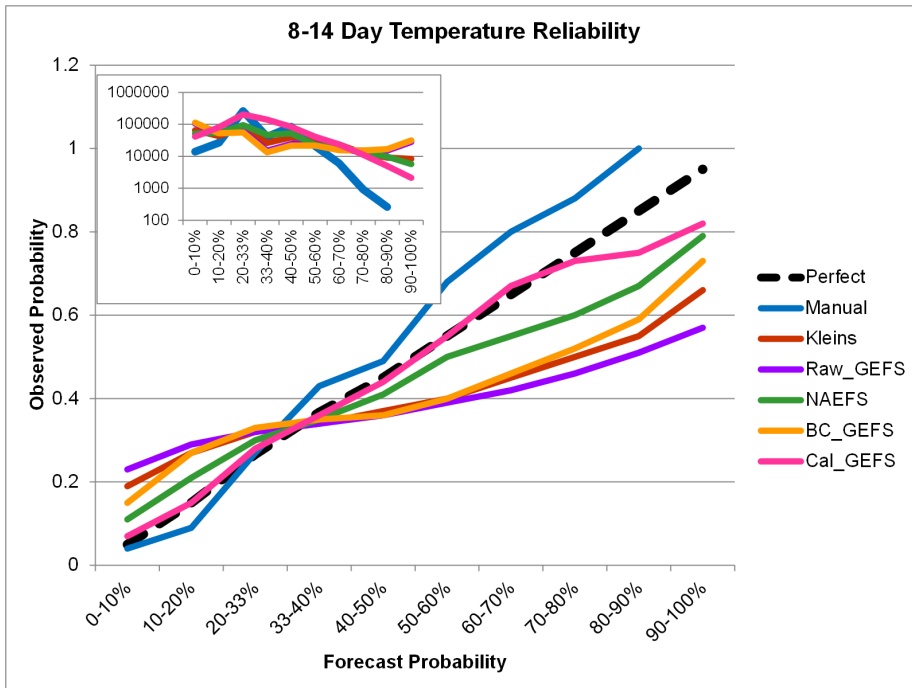
Comparison of reliability of CPC week-2 probability forecasts including official manual forecasts, raw GEFS, bias-corrected GEFS component of NAEFS, NAEFS MME, and reforecast-calibrated GEFS.
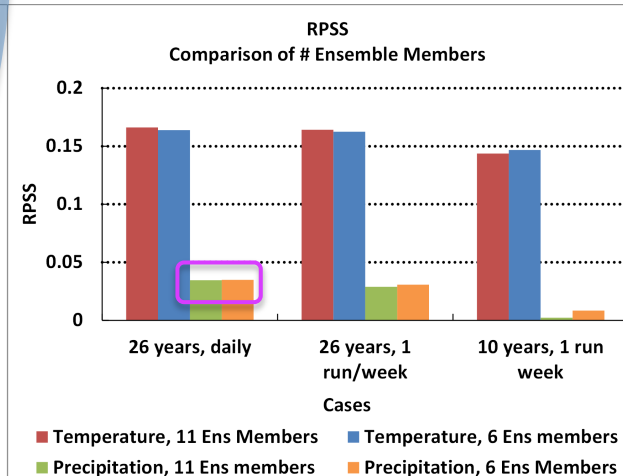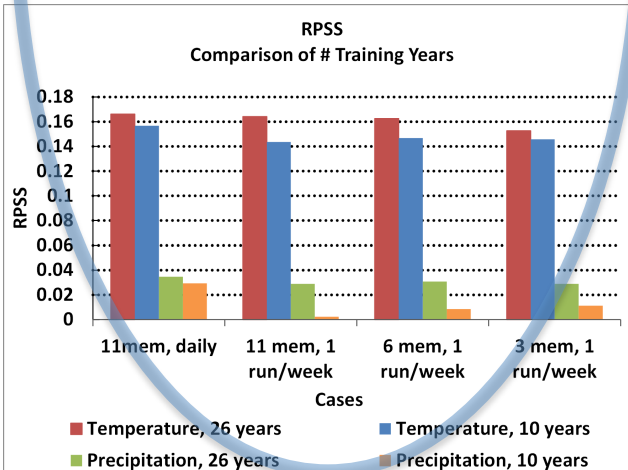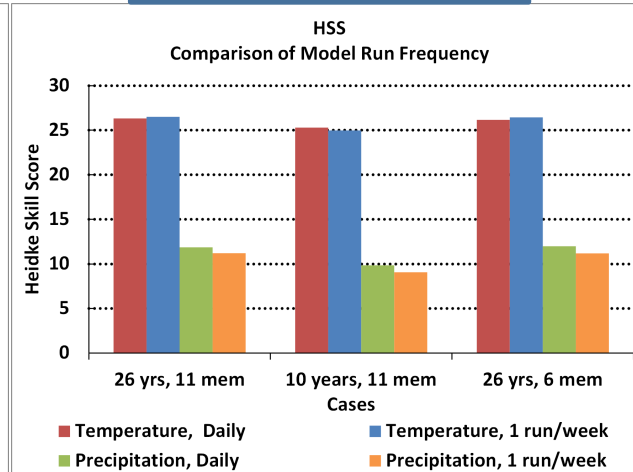
Sampling a smaller reforecast
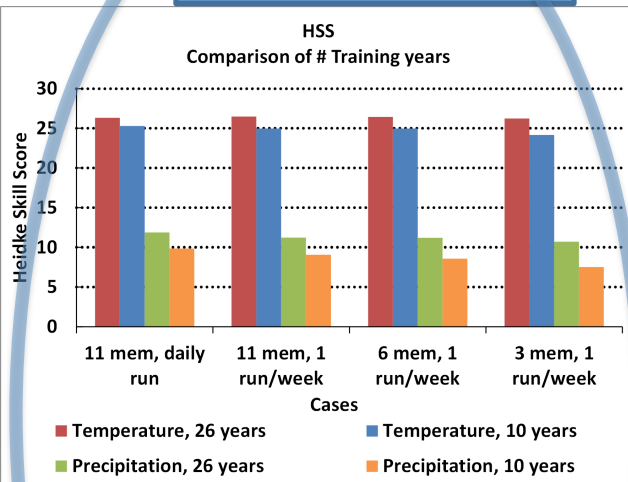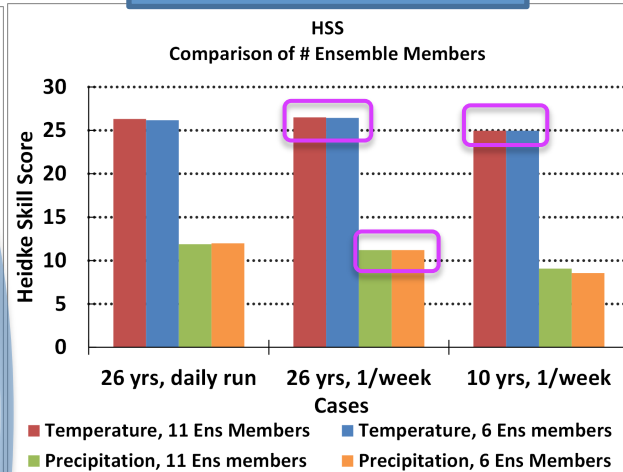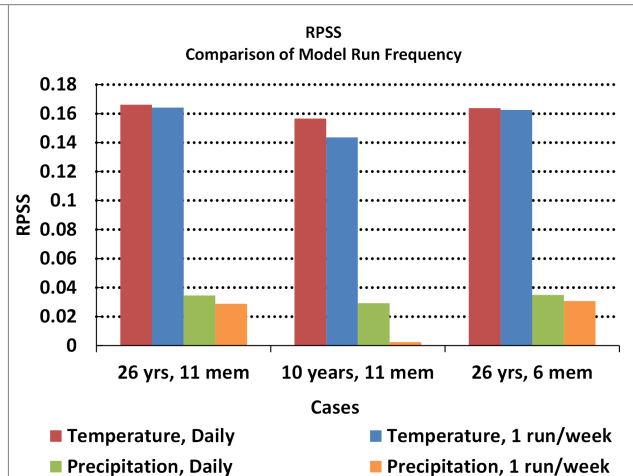
Results of reforecast sample size tests for week-2 temperature and precipitation show reducing the number of years in the reforecast produces the largest drop in skill. Reducing the ensemble size from 11 to 6 members had little or no impact.
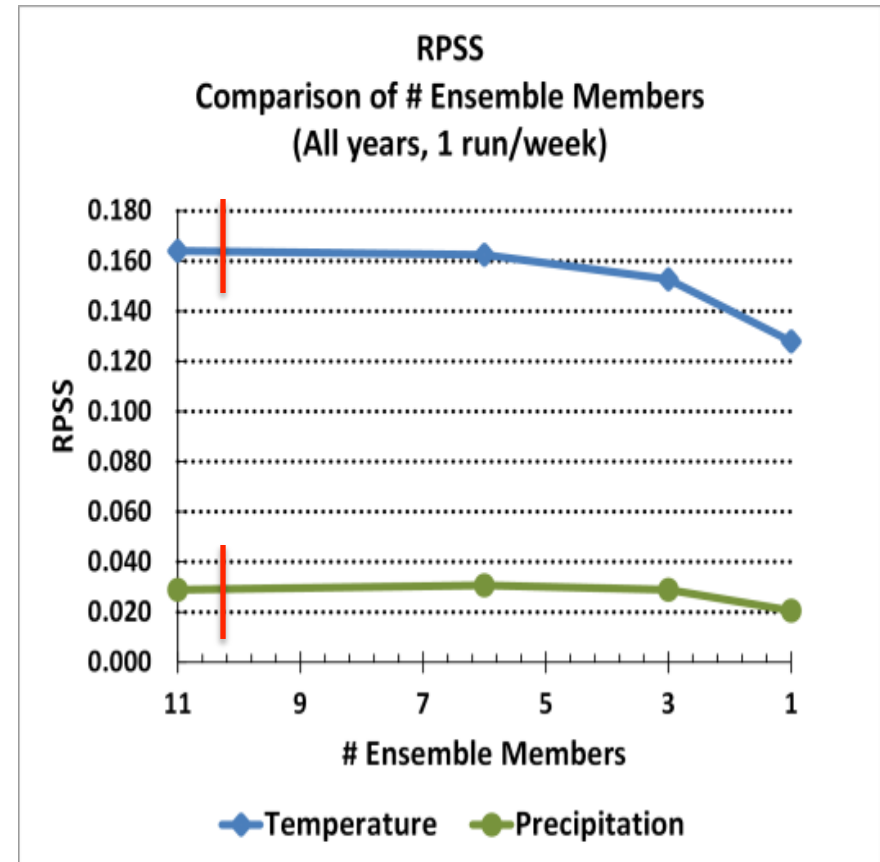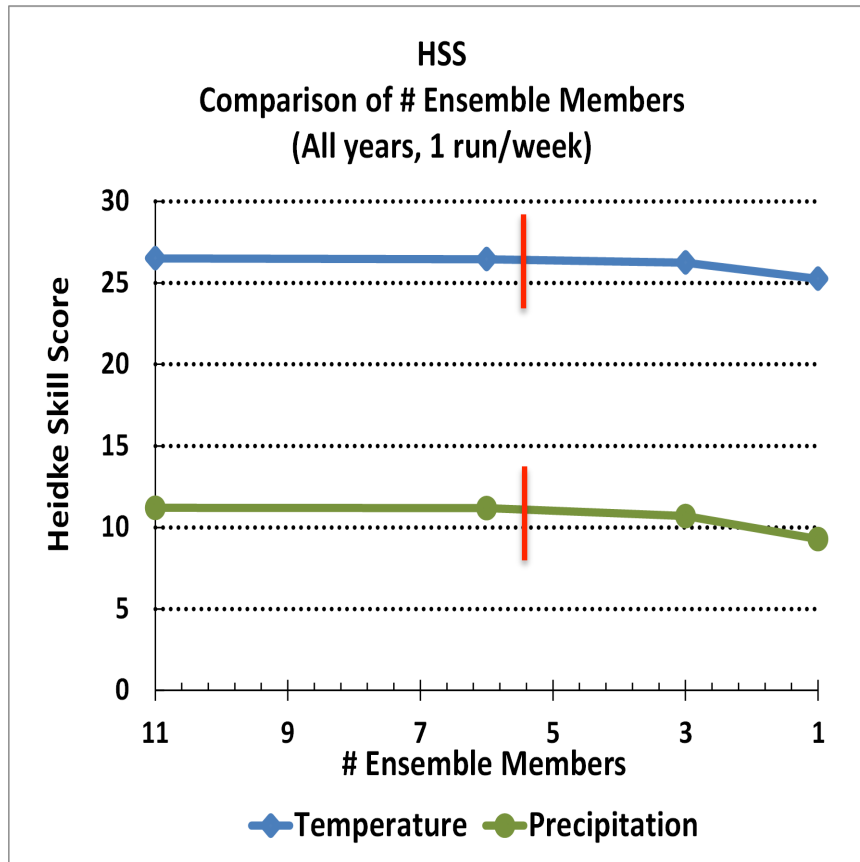
Testing the impact on week-2 temperature and precipitation skill of a reduced number of **years** in reforecast sample
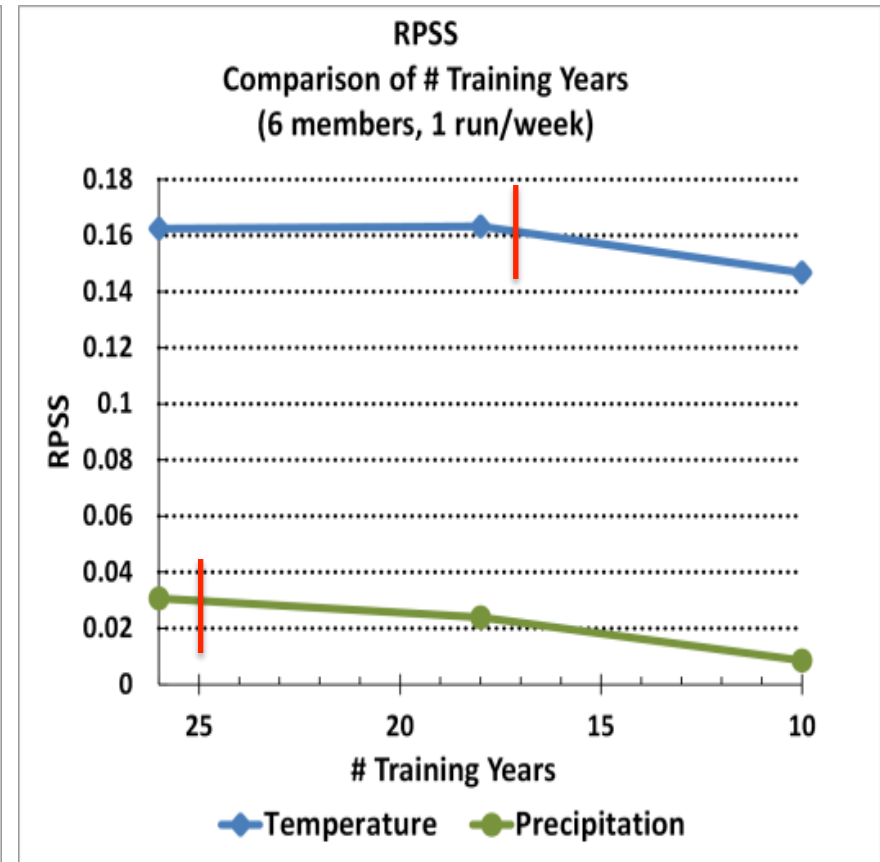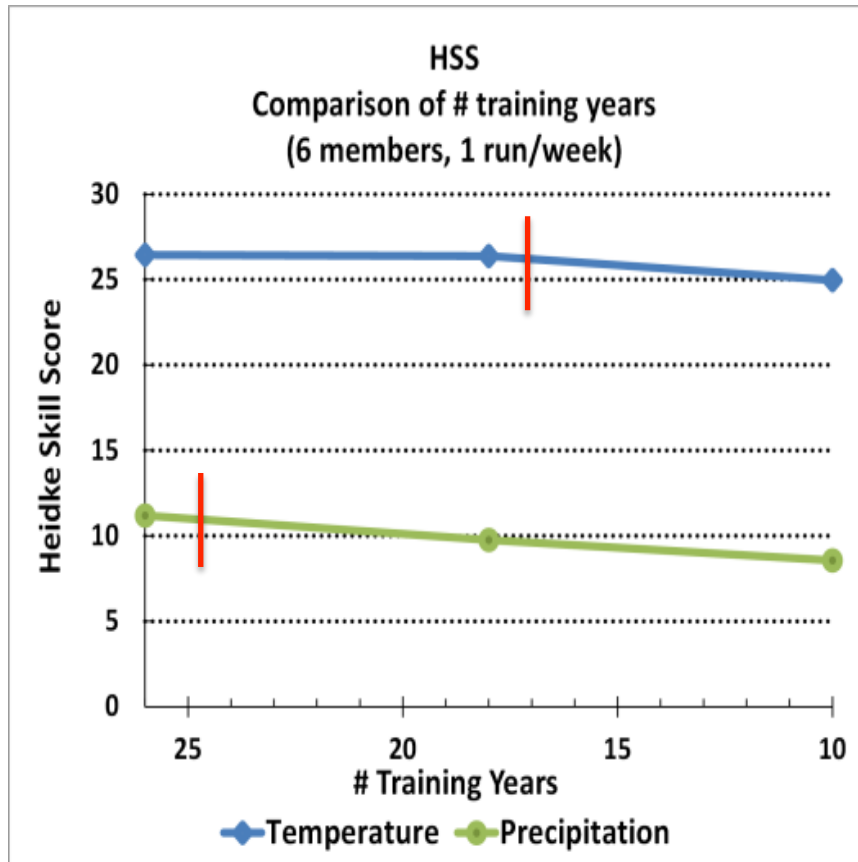
Using 26, 18 and 10 years of reforecasts (1984-2010) to generate statistics for calibration of 2011-2013 GEFS forecasts. Heidke Skill Score (left) and RPSS (right) show *loss of skill for precipitation forecasts with reduction from full 26 years.*
Red lines mark first reduced sample with significantly lower skill.

Testing the impact on week-2 temperature and precipitation skill of a reduced number of **ensemble members** in reforecast sample

Using 11, 6, 3 and 1 ensemble members of reforecasts (1984-2010) to generate statistics for calibration of 2011-2013 GEFS forecasts. Heidke Skill Score (left) and RPSS (right) show *little or no loss in skill with reduction from 11 to 6 ensemble members.*

Testing the impact on week-2 temperature and precipitation skill of a reduced frequency of runs from daily to weekly
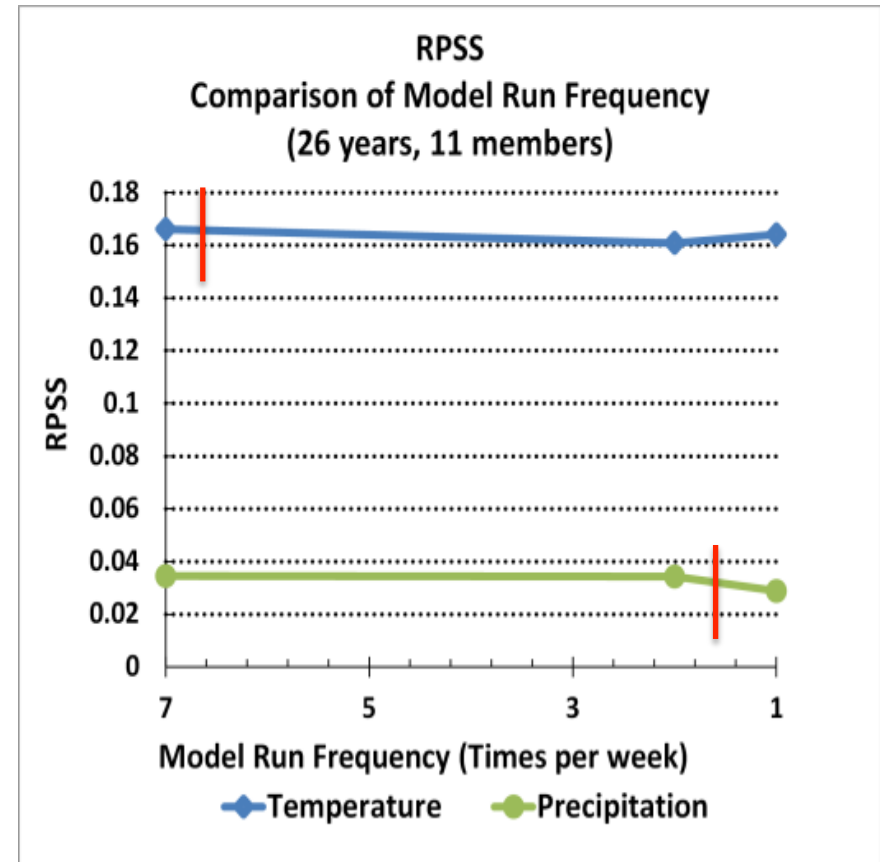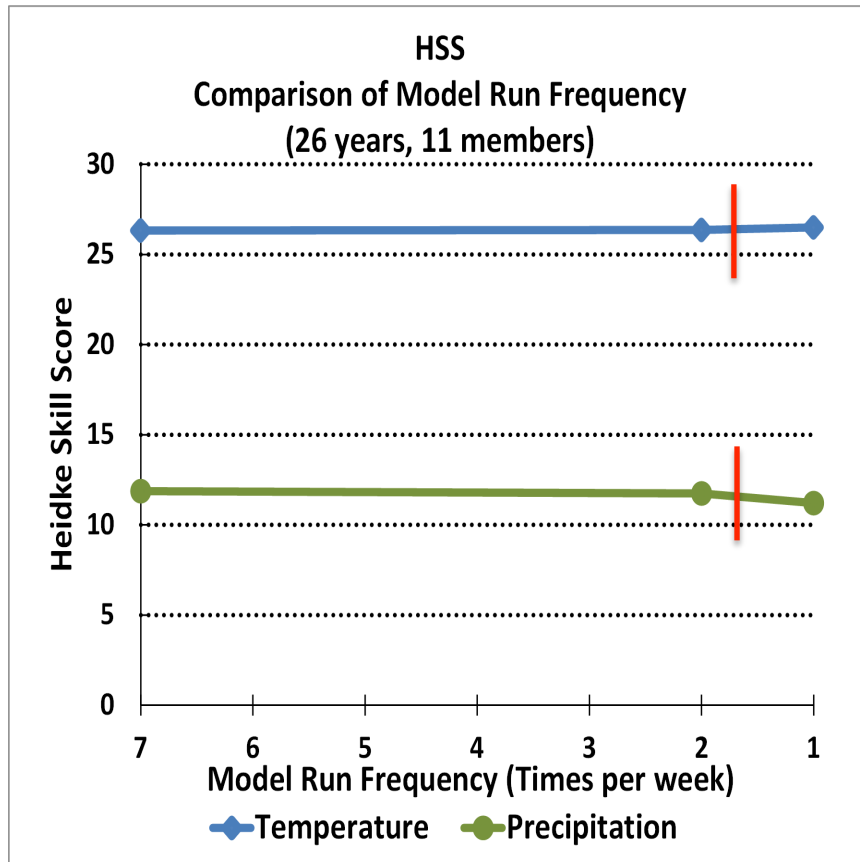
Using 7, 2 and 1 run per week (1984-2010) of reforecasts to generate statistics for calibration of 2011-2013 GEFS forecasts. Heidke Skill Score (left) and RPSS (right) show *little or no loss in skill with reduction from daily to twice-weekly runs.*

Using Ensembles for Predictability Estimates

What do we mean by predictability?

- The **ensemble mean** is an estimate of the predictable component of the future state, also known as *signal S*

$$x_t = \langle x \rangle_t + x'_t$$

- The **ensemble spread** about the mean is an estimate of the unpredictable *noise*, **N**

$$N_t = \sqrt{\frac{1}{n} \sum (x'_t)^2}$$

- By these estimates, signal and noise are dependent on the initial state of the climate system at time, *t*.

Shifts in the probability distribution indicate the strength of the signal
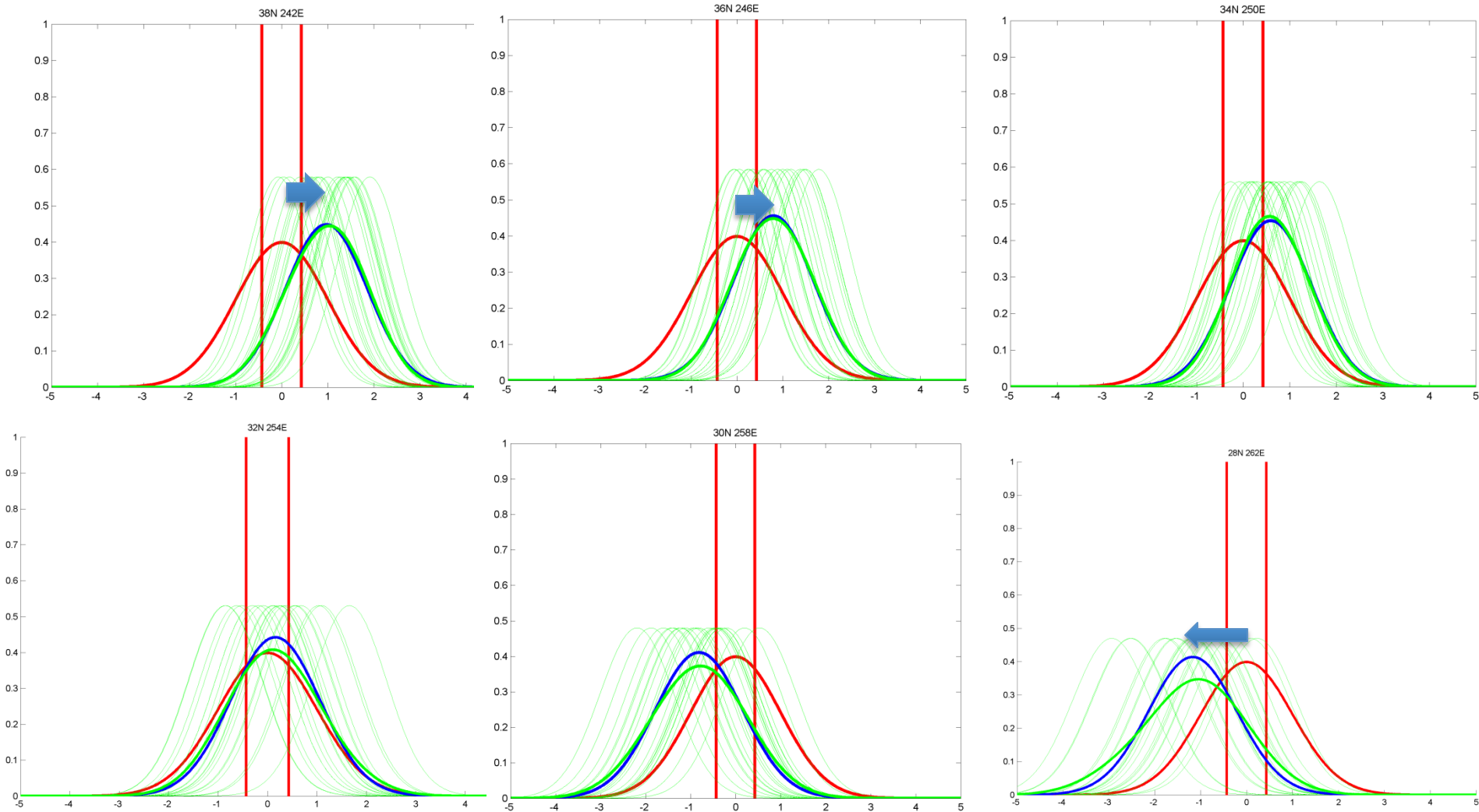
Spread of ensemble members determine the uncertainty or noise
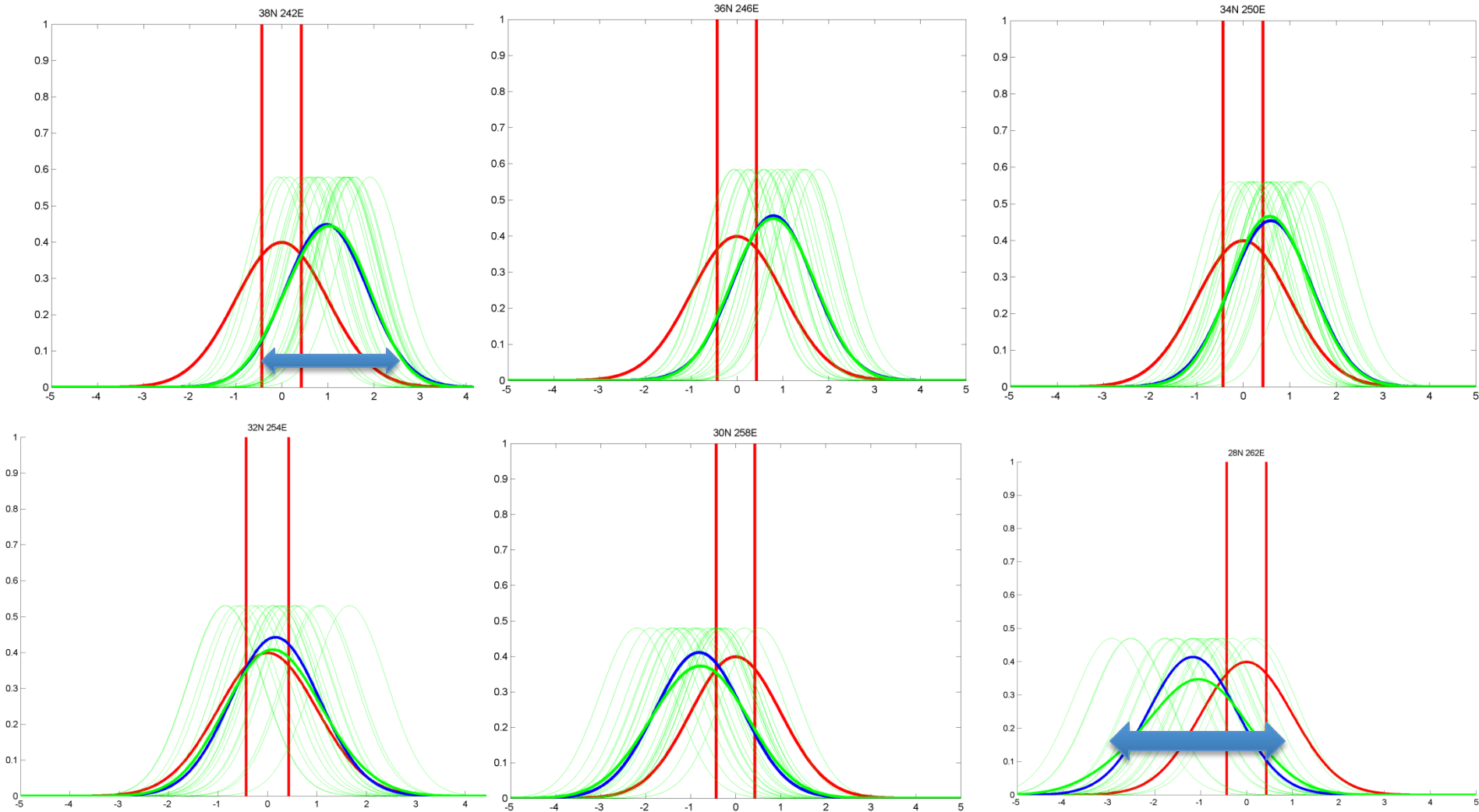
Using Ensembles for Predictability Estimates(2)

- A measure of the predictability of the future climate state <u>specific to a subset of initial conditions </u> is the *signal-to-noise ratio*

$$SNR_{predictable} = \sqrt{\frac{\overline{S}^2}{\overline{N}^2}} \qquad \overline{S}_{predictable} = \sqrt{\sum_t \langle x_t \rangle^2}$$

i.e. The variability of the ensemble mean relative to the variability of the ensemble members about the mean.

- The variance of the ensemble mean and members can be used as an estimate of the potential skill of forecasts

$$R_{predictable} = \sqrt{\frac{\overline{S}^2}{\overline{S}^2 + \overline{N}^2}}$$

Potential Correlation as a function of the signal-to-noise ratio, and ensemble size (blue:1 member, black: unlimited ensemble)

Ensemble Mean Correlation
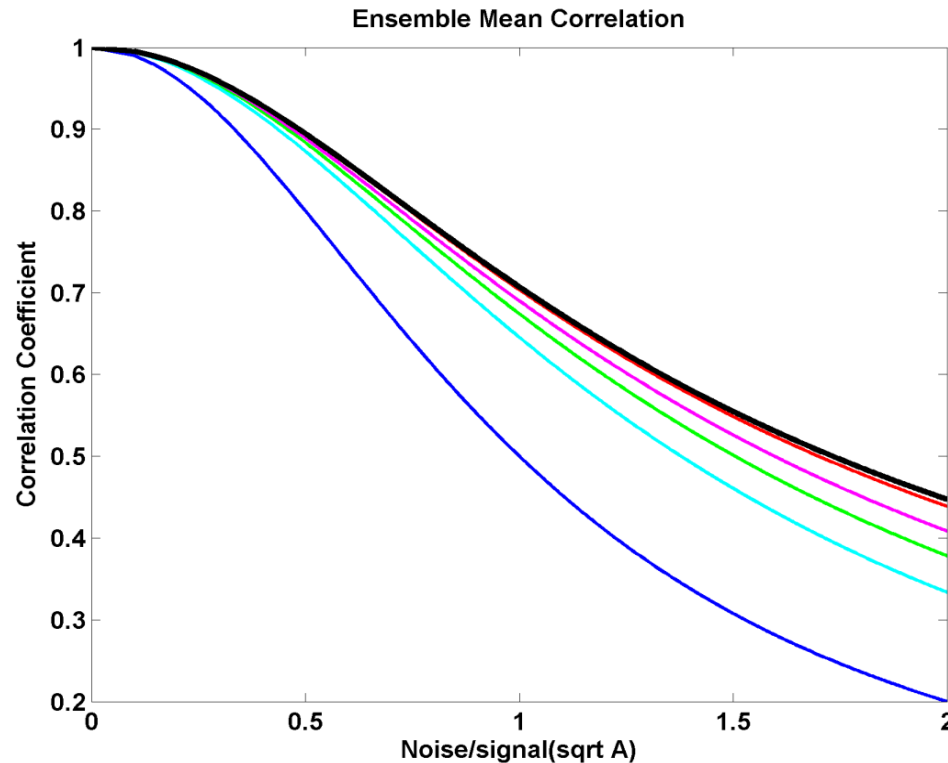
- This places an upper limit on predictability as a function of the SNR   (*courtesy of Emily Riddle*)

Changes in predictability

- We can consider the linear changes over a number of years

- Linear trends can be calculated for signal (S), noise (N), SNR, and the potential predictability indicated by the prediction of the correlation by the model (R).

1) *It is a reasonable assumption that the predictability of particular characteristics of the climate system may change as the background climate state changes*

2) *By this measure, predictability may also change with changing observational networks*

Why study predictability?

1. To evaluate the performance of forecast systems for different conditions and initial states

2. To make statistical corrections to dynamical forecasts

3. To describe and understand the innate uncertainty in the chaotic climate system

Using NOAA-ESRL retrospective forecasts

- Current operational NCEP Global Ensemble Forecast System (GEFS) as of February 2012
- 10 perturbation ensemble members + control (best estimate of initial conditions)
- Week-2, 8-14 day lead forecasts of 2-meter air temperature
- 1985-2010
- Initial conditions from Climate Forecast System Reanalysis (CFSR)
- CFSR used as observations
- 25 year record includes climate change + changes in observational network

## Using retrospective forecasts for bias correction

- Determining the systematic error generally assumes the statistics are stationary over reforecast data set.

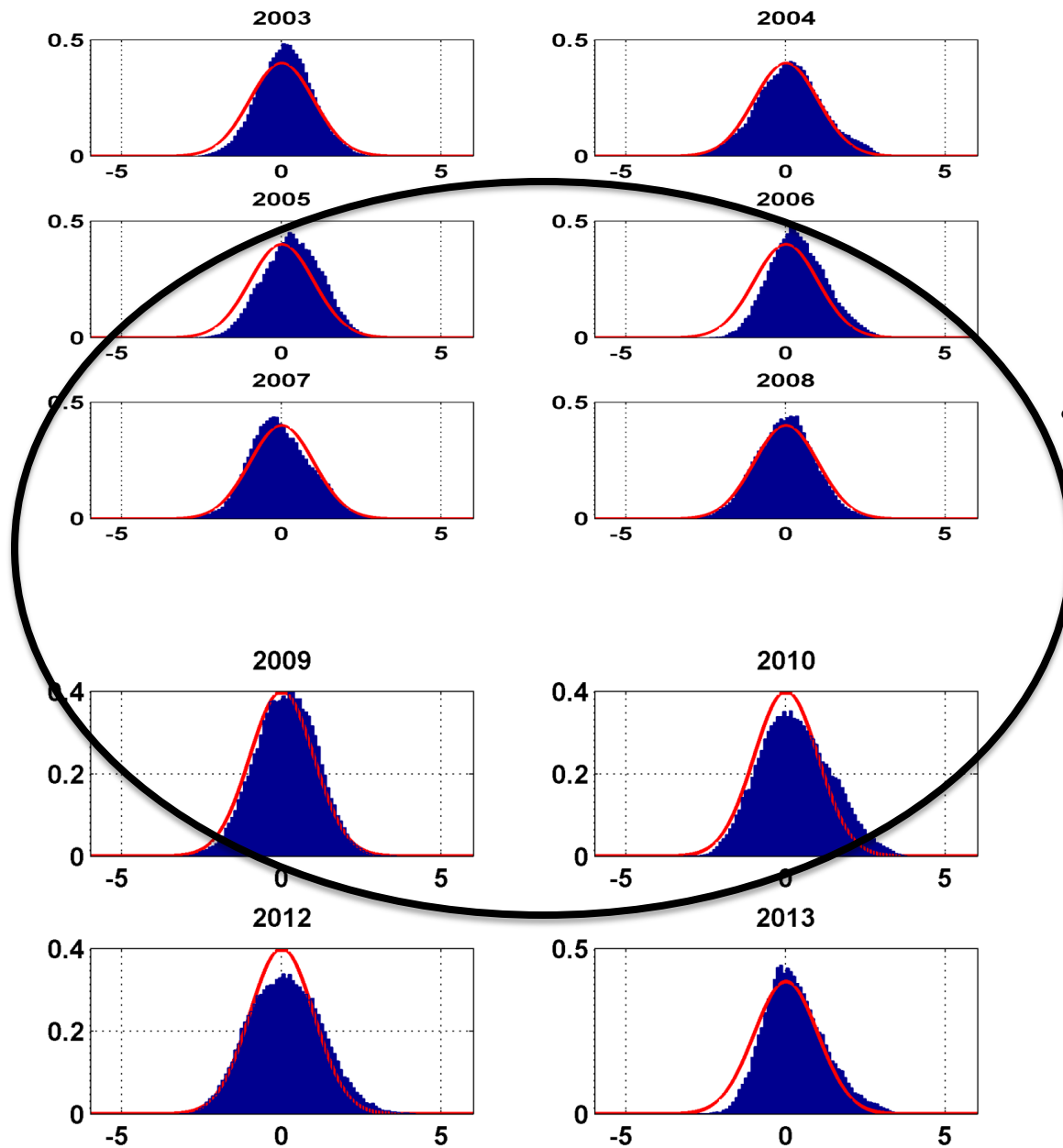- **Are the statistics stationary?**
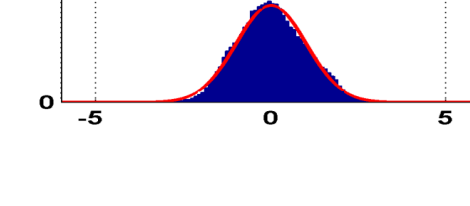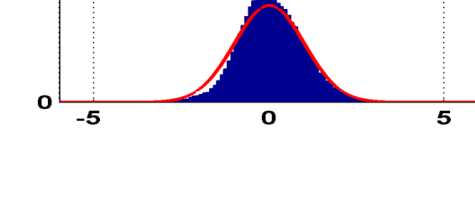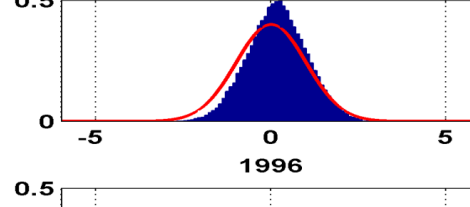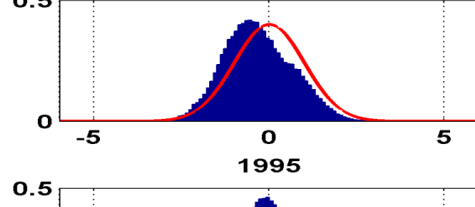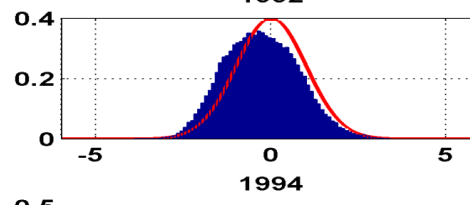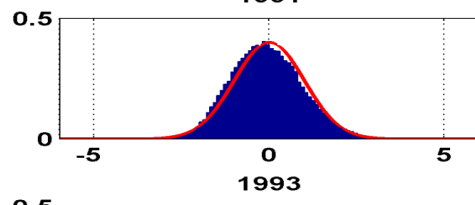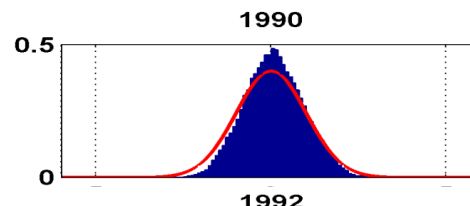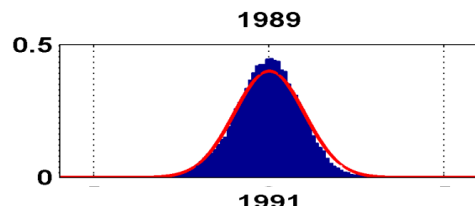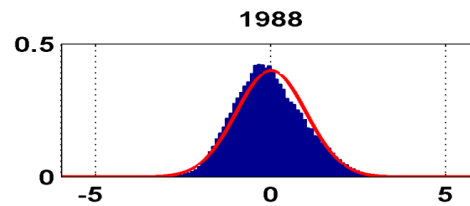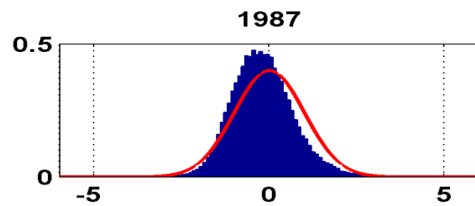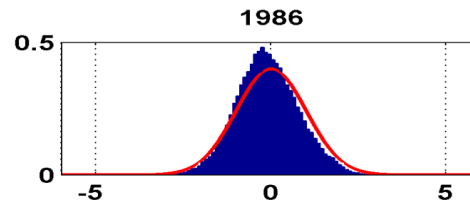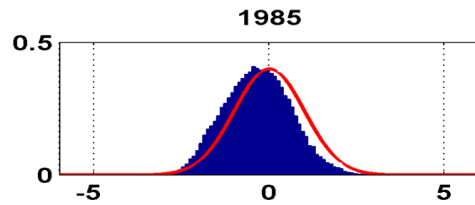
- Weekly mean forecast temperatures over the U.S. by year
- 25-year climatology in red
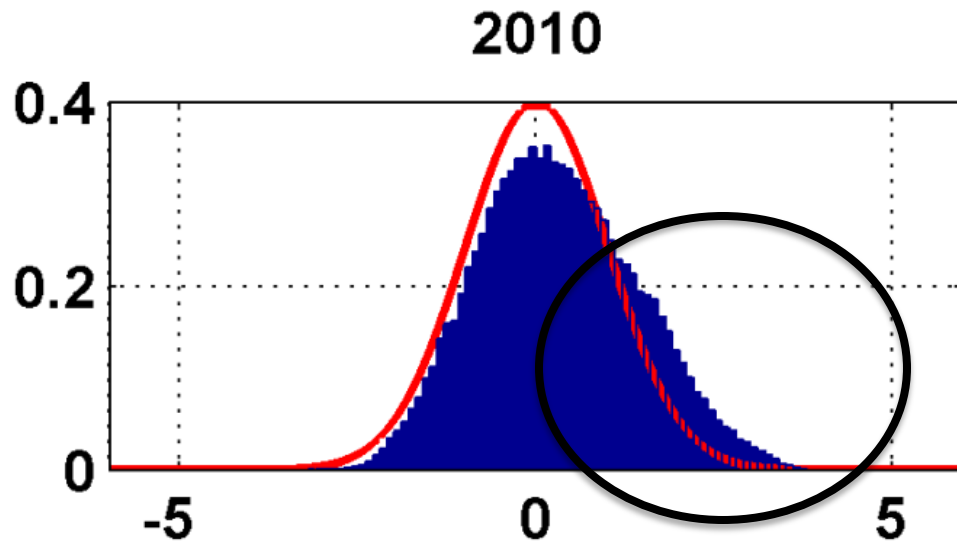- PDF shifting each year
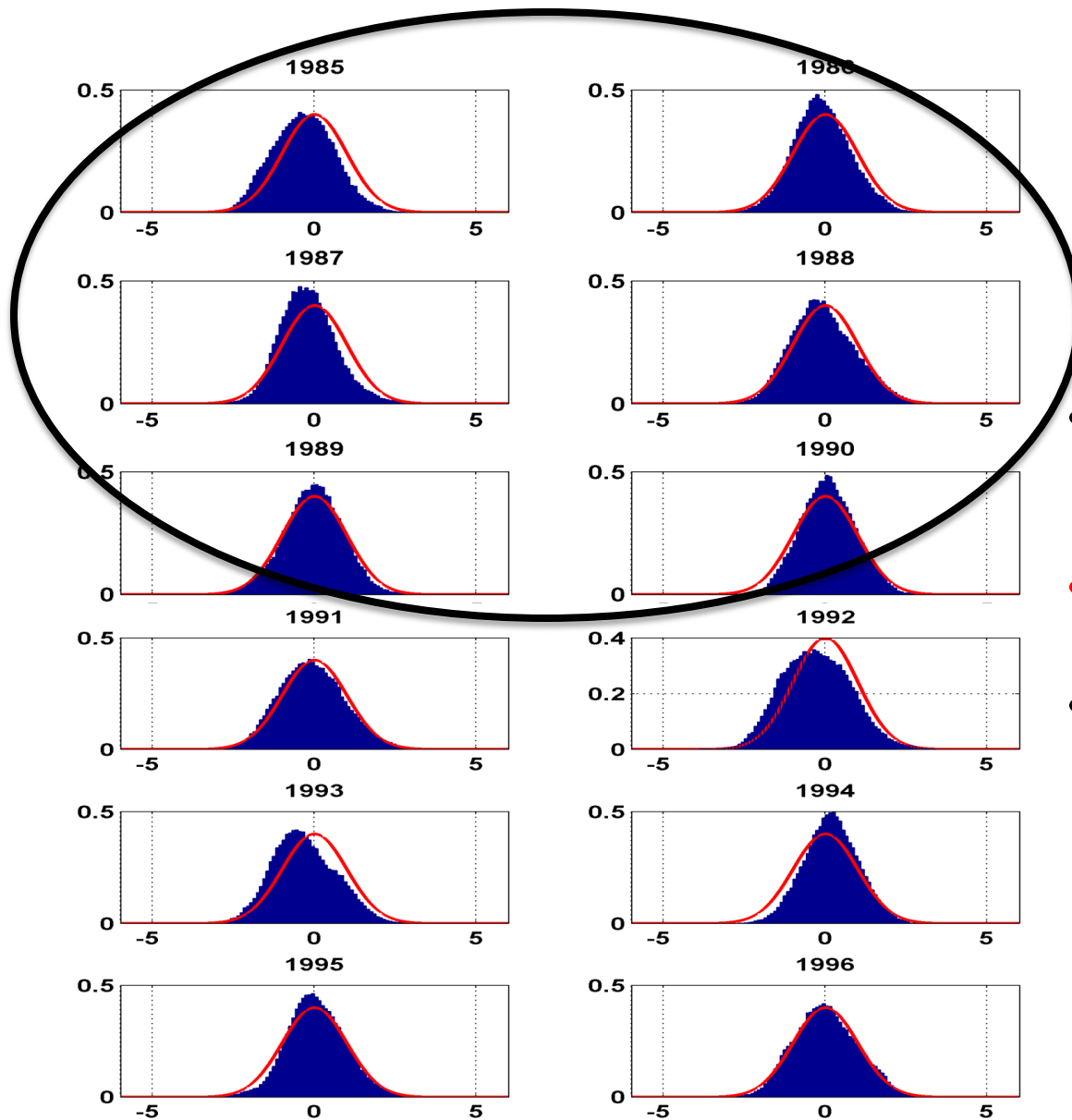
- Weekly mean temperatures over the U.S. by year
- 25-year climatology in red
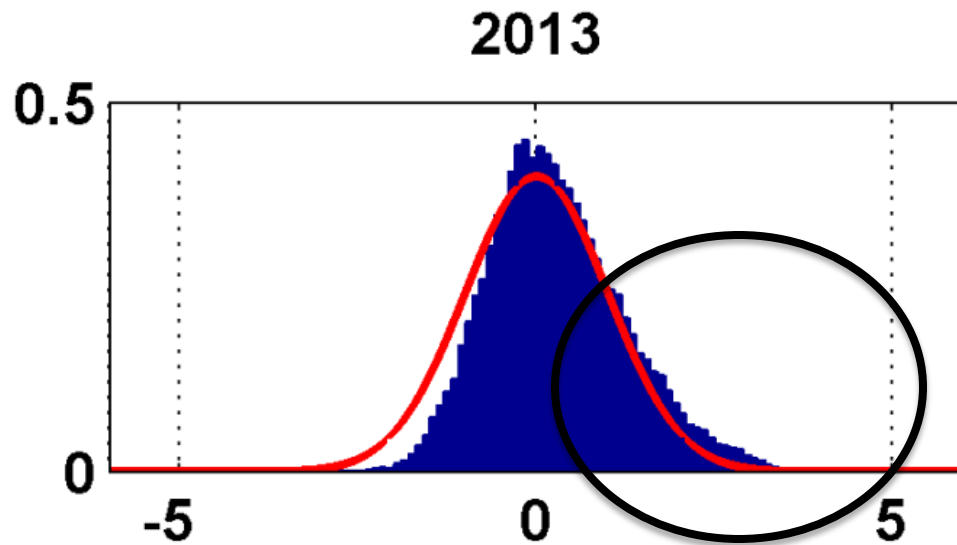- PDF shifting each year
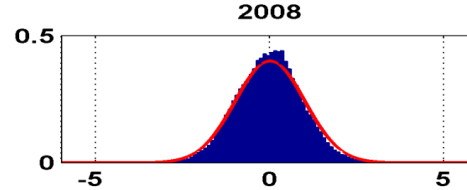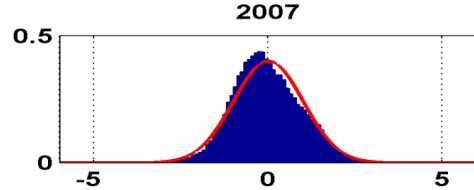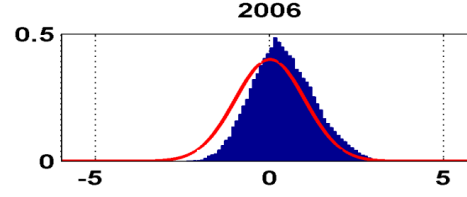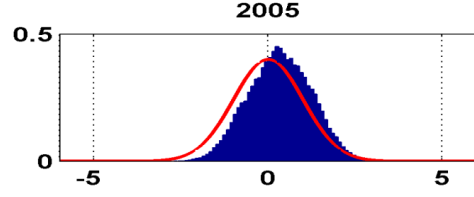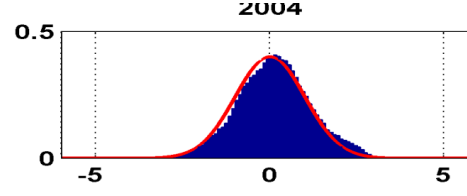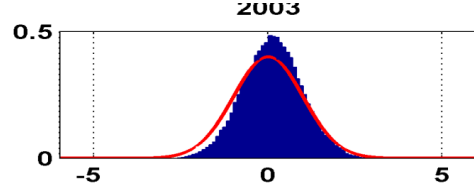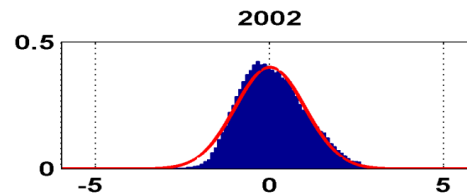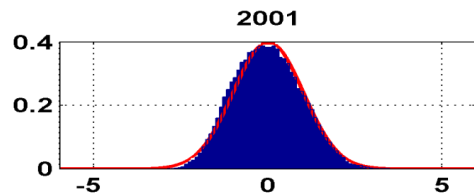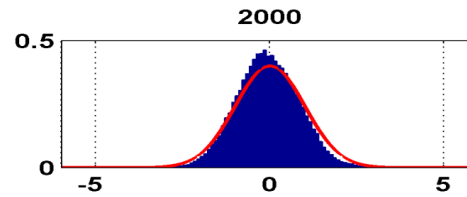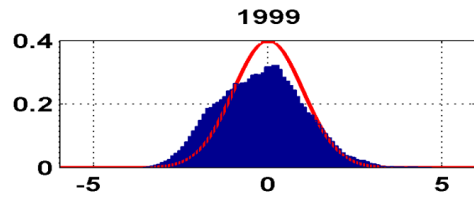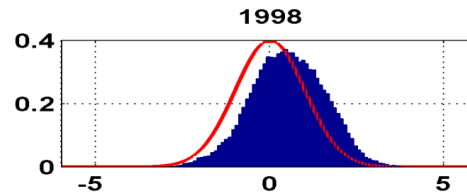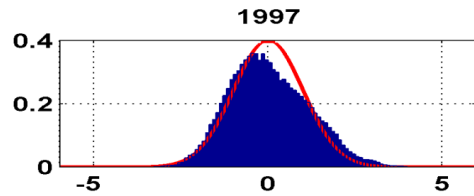
- Multi-decadal temperature trends apparent in intraseasonal forecasts
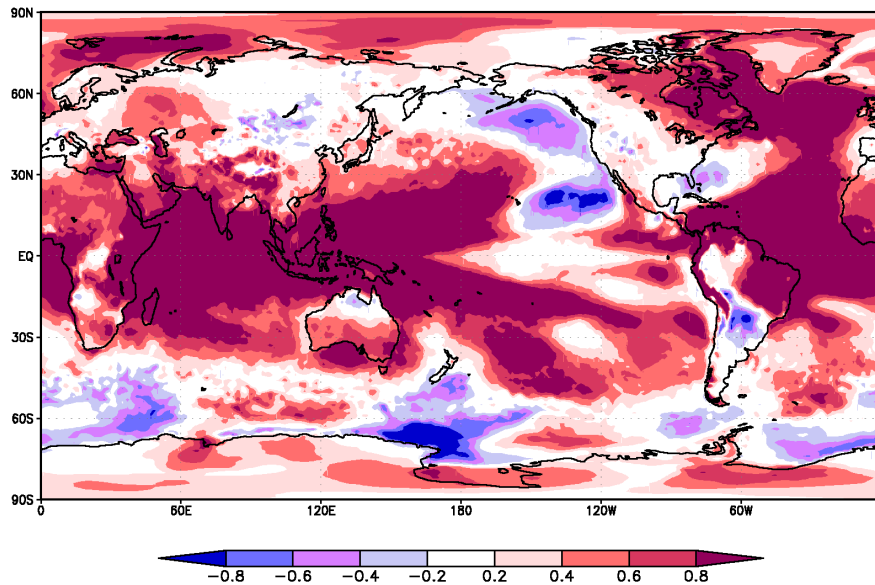
Shift towards warmer temperatures in more recent years

2010

2013

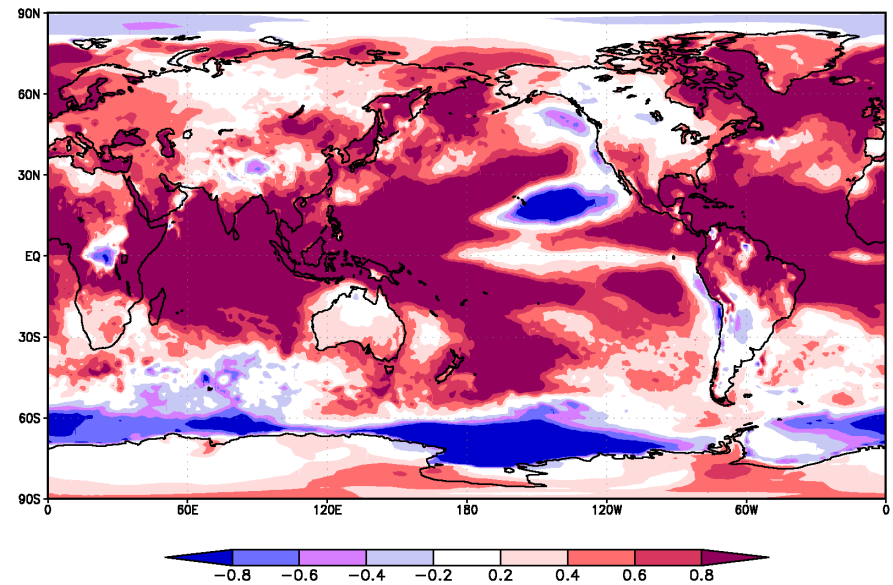- Increase in the frequency of extremes

CFSR standardized (weekly) linear temperature trend for 25-year period (1985 to 2010)
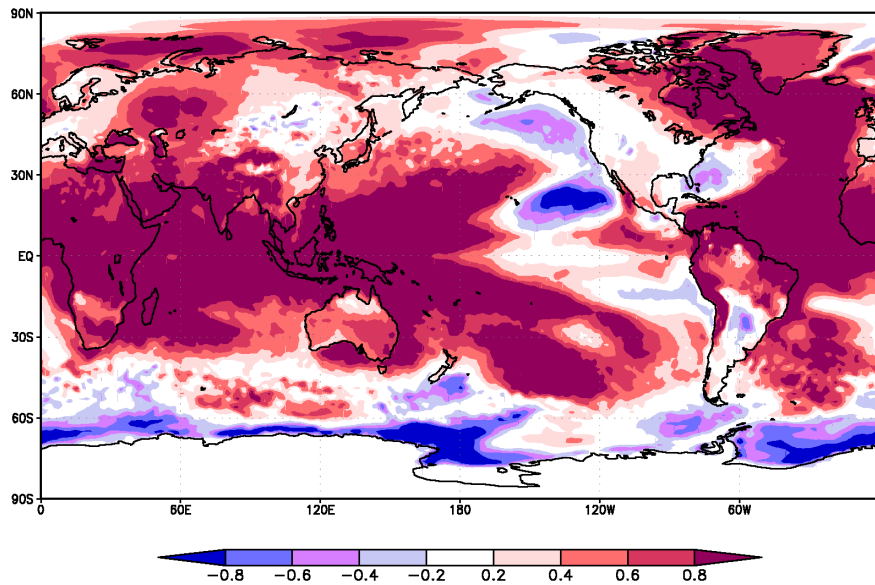
November to March
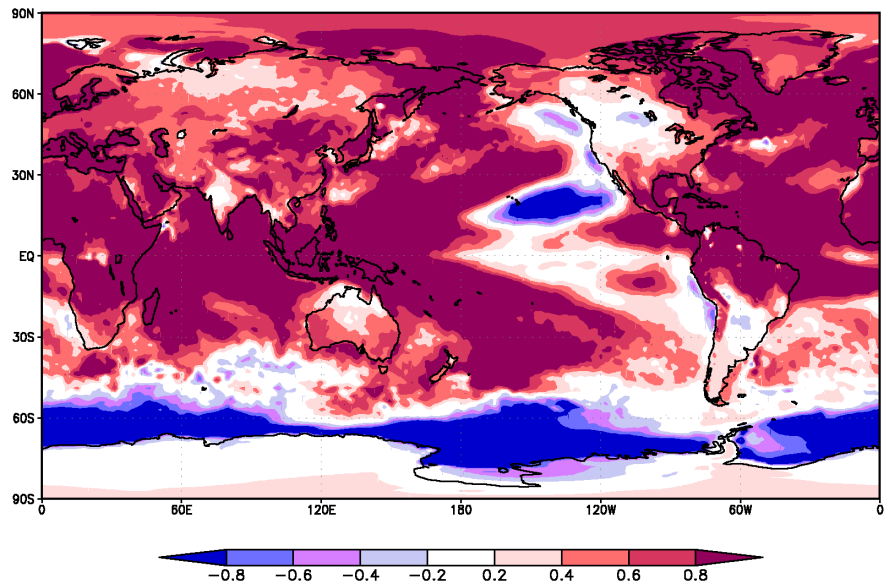
May to September

- Temperature trends are significant fraction of weekly timescale variability

- Large areas exceed 0.5 standard deviations

Model week-2 standardized linear temperature trend for 25-year period (1985 to 2010)
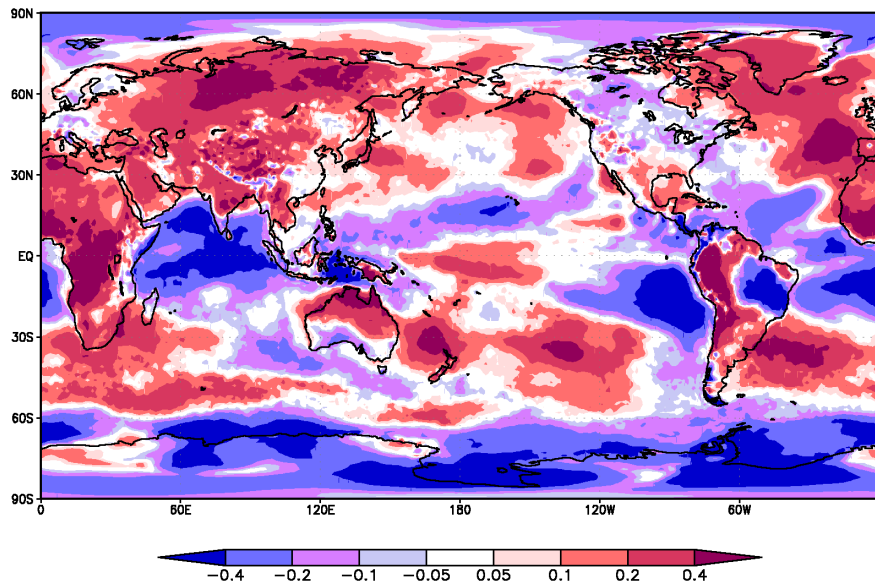
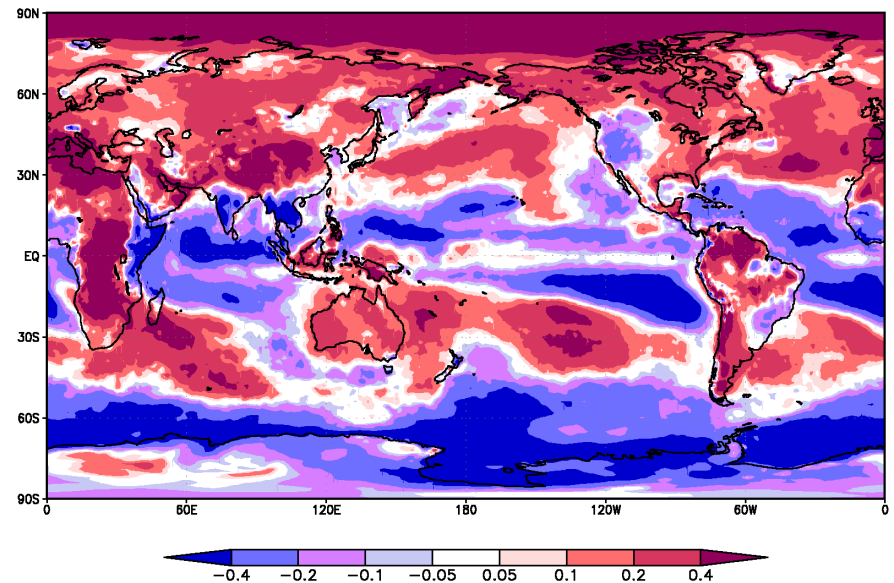November to March   May to September

- Decadal changes in the model climate are similar to initialization reanalysis

- Many significant differences globally (bias in the trends)

Standardized trends in **bias** of model week-2 forecasts for 25-year period (1985 to 2010)

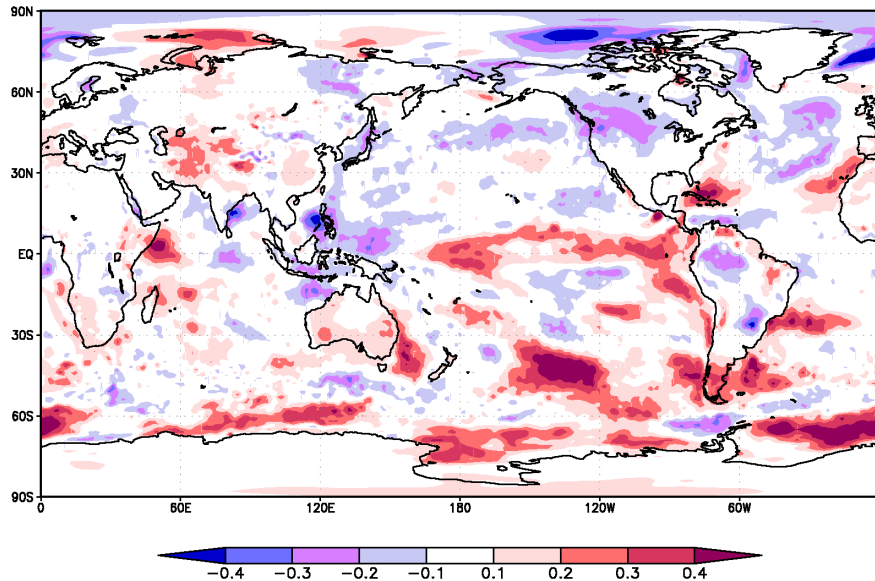November to March   May to September
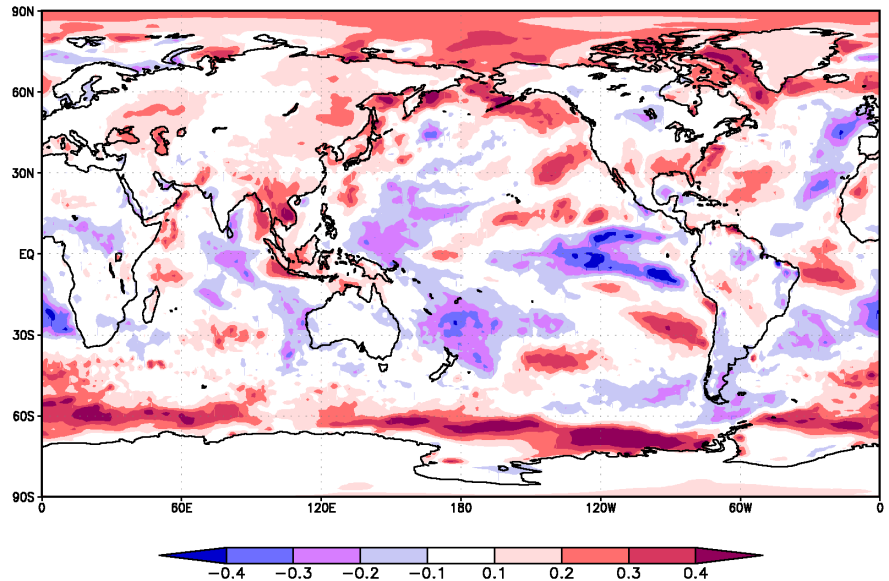

- Bias changes are significant fraction of variability

- Changes in bias over time are highly variable by location and season

Model week-2 standardized trend in **signal** for 25-year period (1985 to 2010)
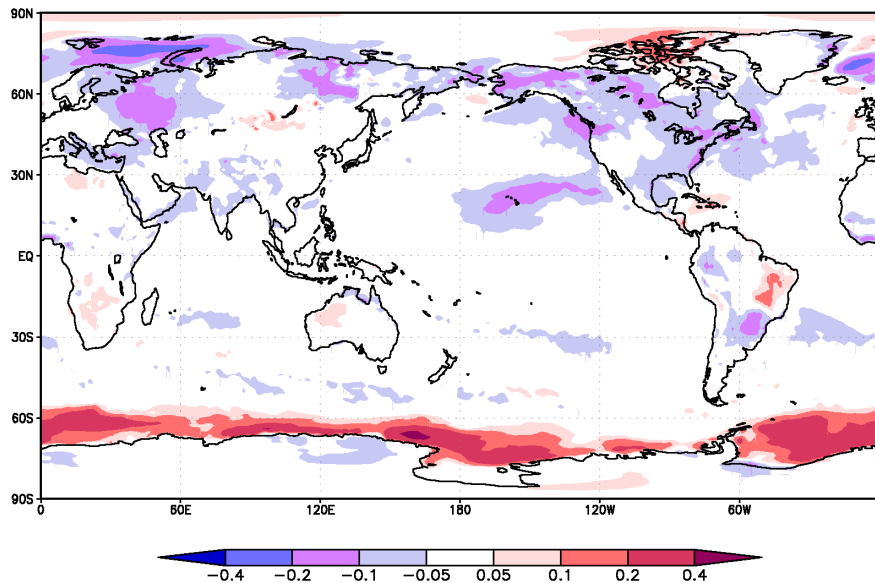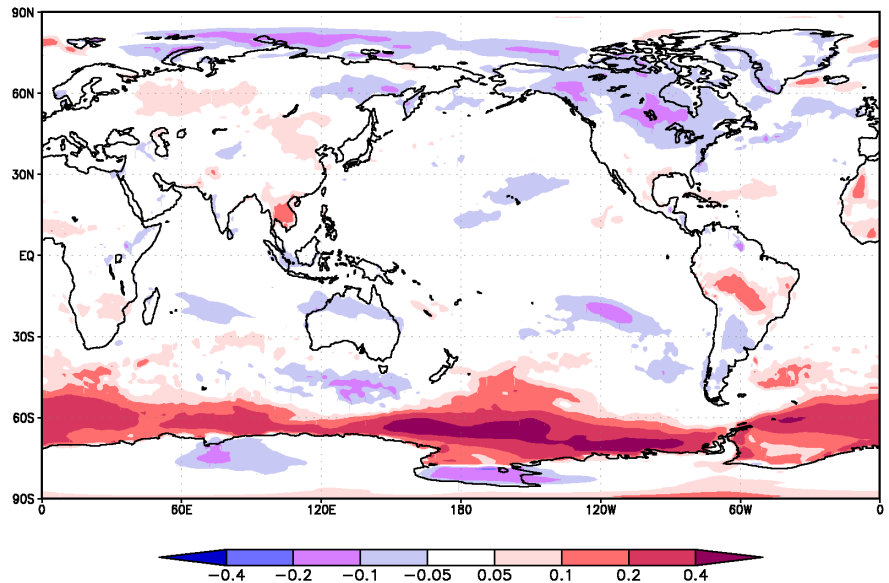
November to March

May to September

- Decadal changes in signal strength are variable by location and season

- Is this changes in observations? Decadal variability of the climate system? Chance changes in climate variability?

Model week-2 standardized trend in **noise** for 25-year period (1985 to 2010)
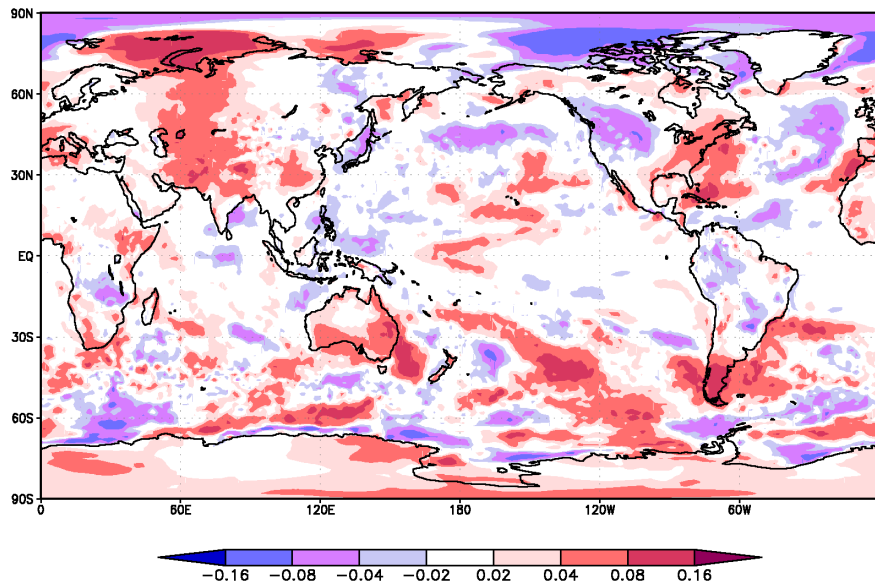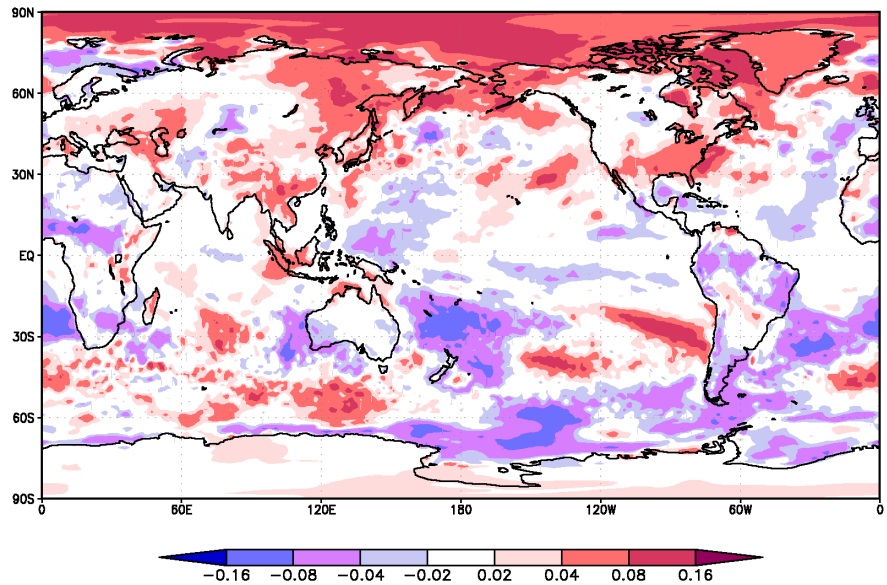
November to March

May to September

- Negative trend in the climate noise in North America. (Better initial conditions?)
- Weak trends in noise overall

Trend in *model-predicted* **potential correlation (predictability)** for 25-year period (1985 to 2010)

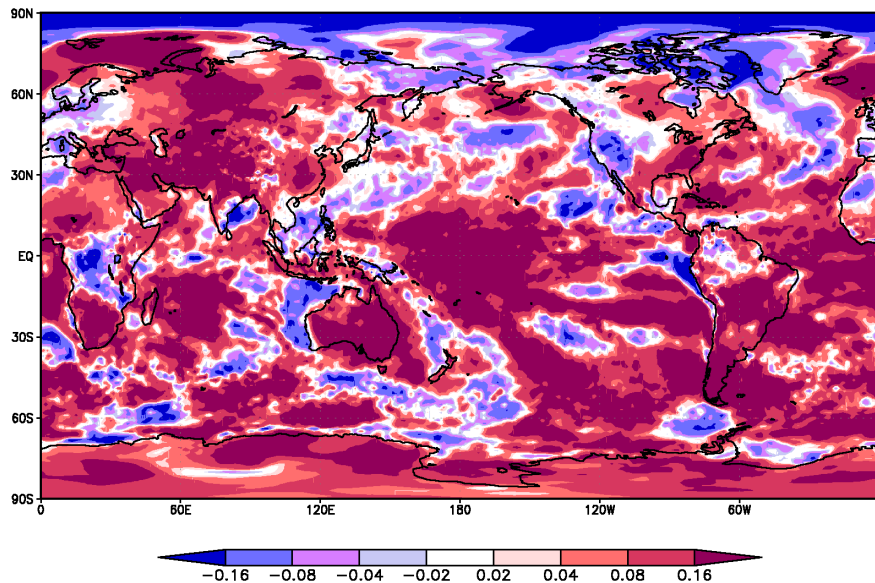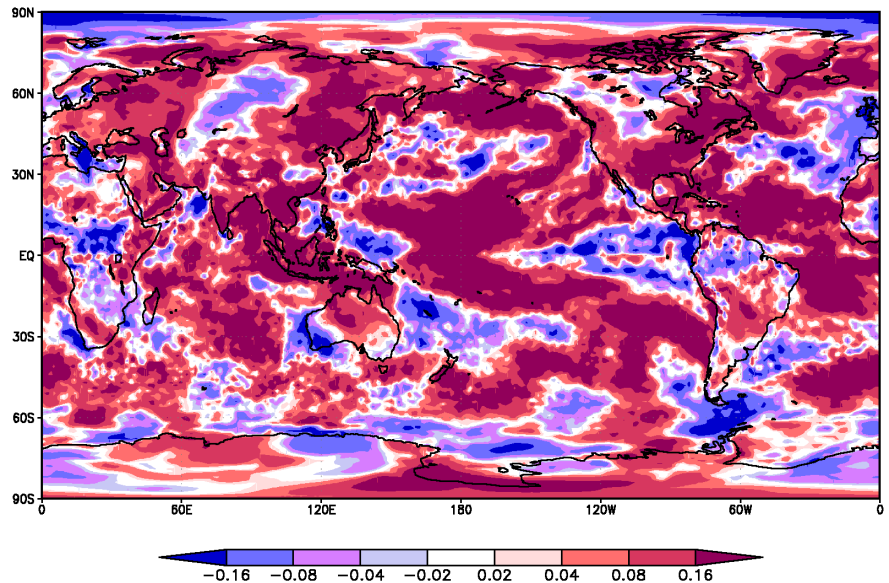November to March   May to September

- Potential positive trend in skill in the Summer hemisphere

Model week-2 forecast **observed correlation trend** for 25-year period (1985 to 2010)
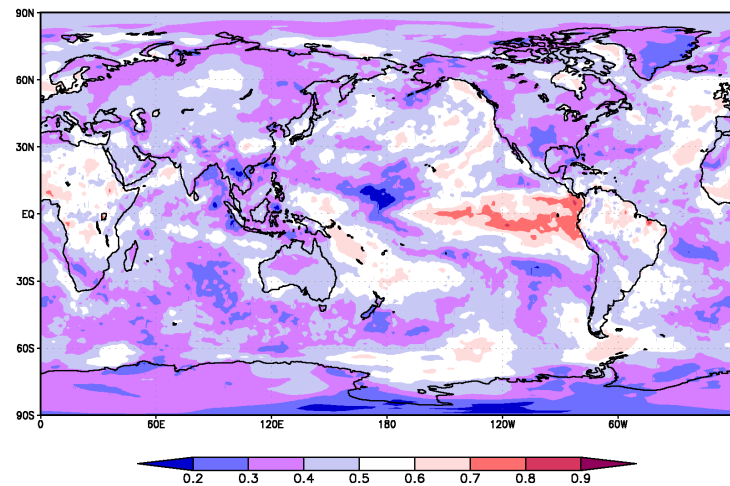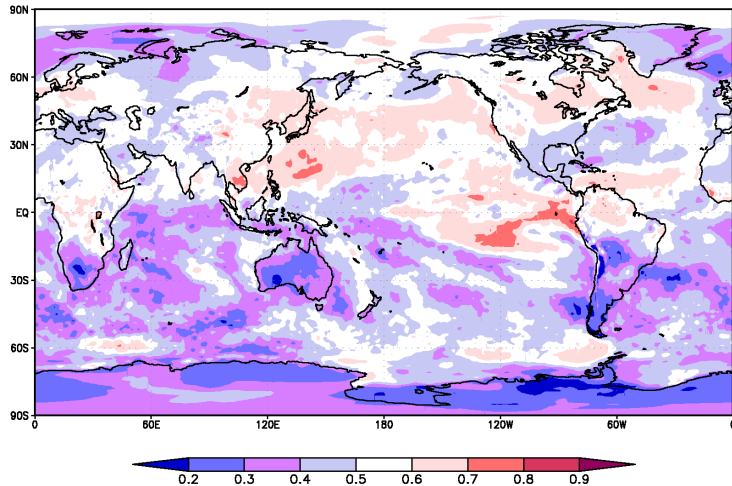
November to March

May to September

- Some similarities in spatial pattern of trends in correlation to observations compared to the model-only potential correlation trends (previous slide)
- Lower observed correlation relative to potential correlation results in greater increase, as the signal increases in magnitude
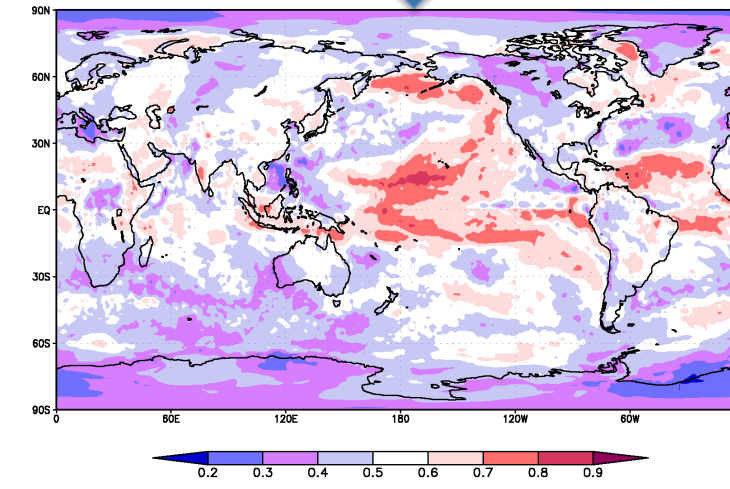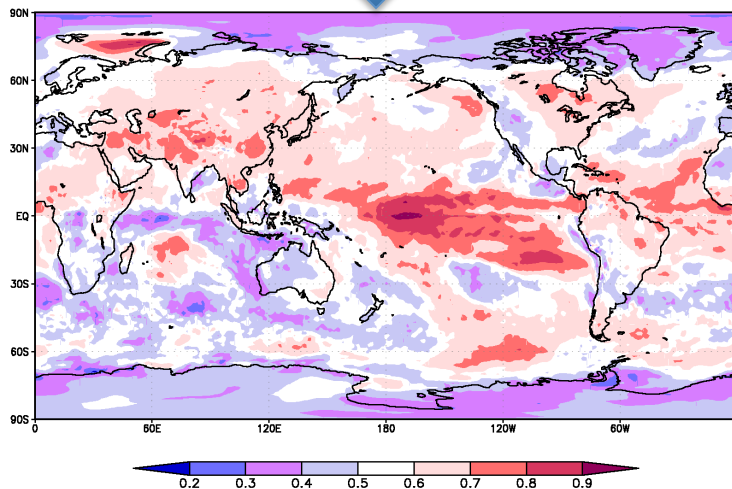
Forecast-analysis **correlation** for 1985 and 2010 accounting for observed trend

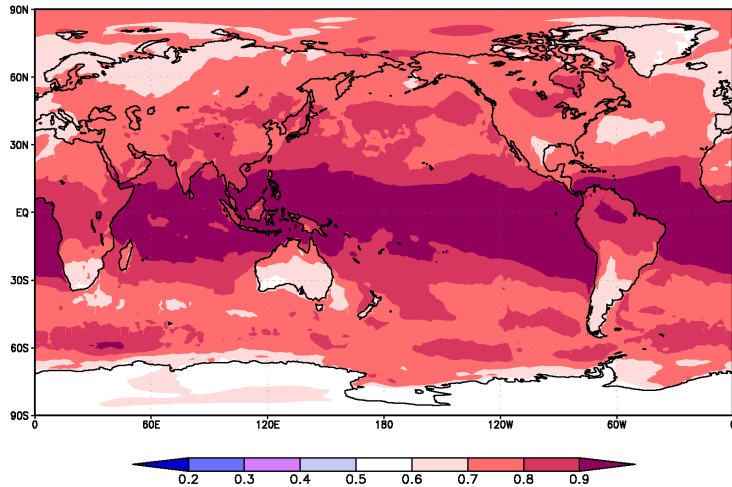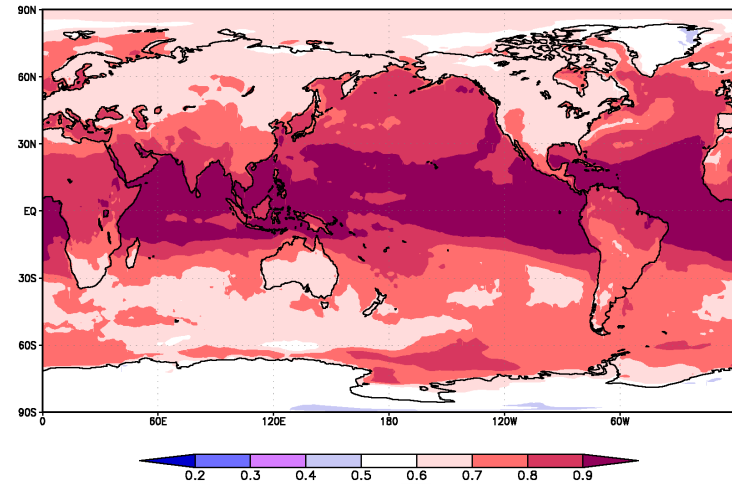## November to March

May to September

1985

2011

Model week-2 **potential correlation (predictability)** for 1985 and 2010 accounting for trend in signal
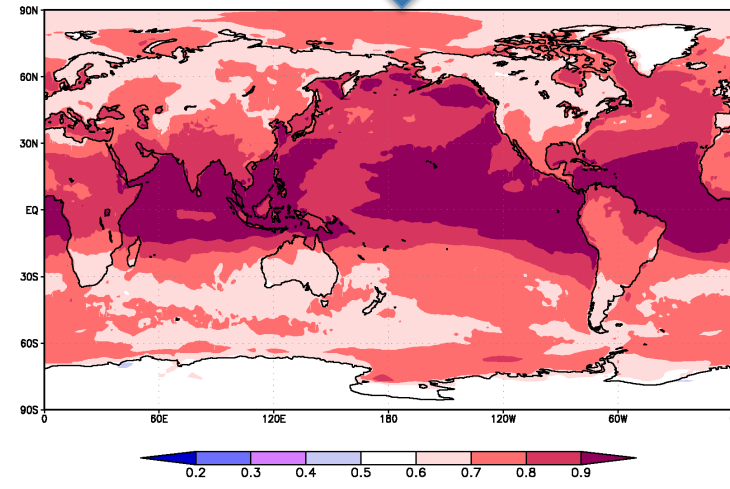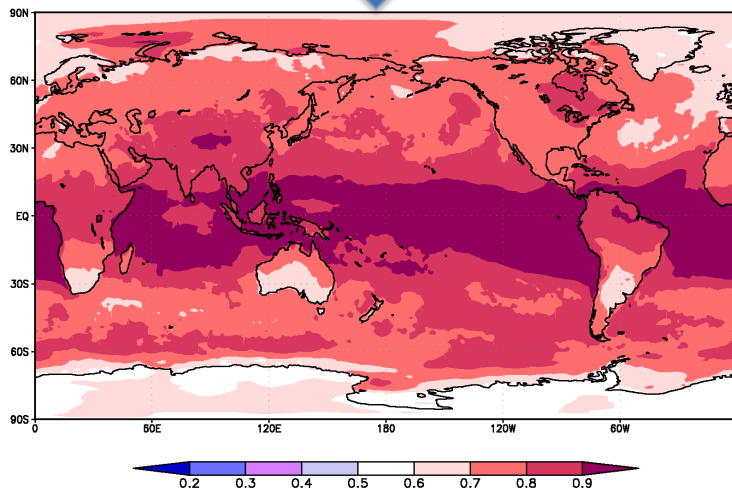
November to March

May to September

1985

2011

Is model bias (right) changing with changing background climate state (Winter, Dec-Jan)
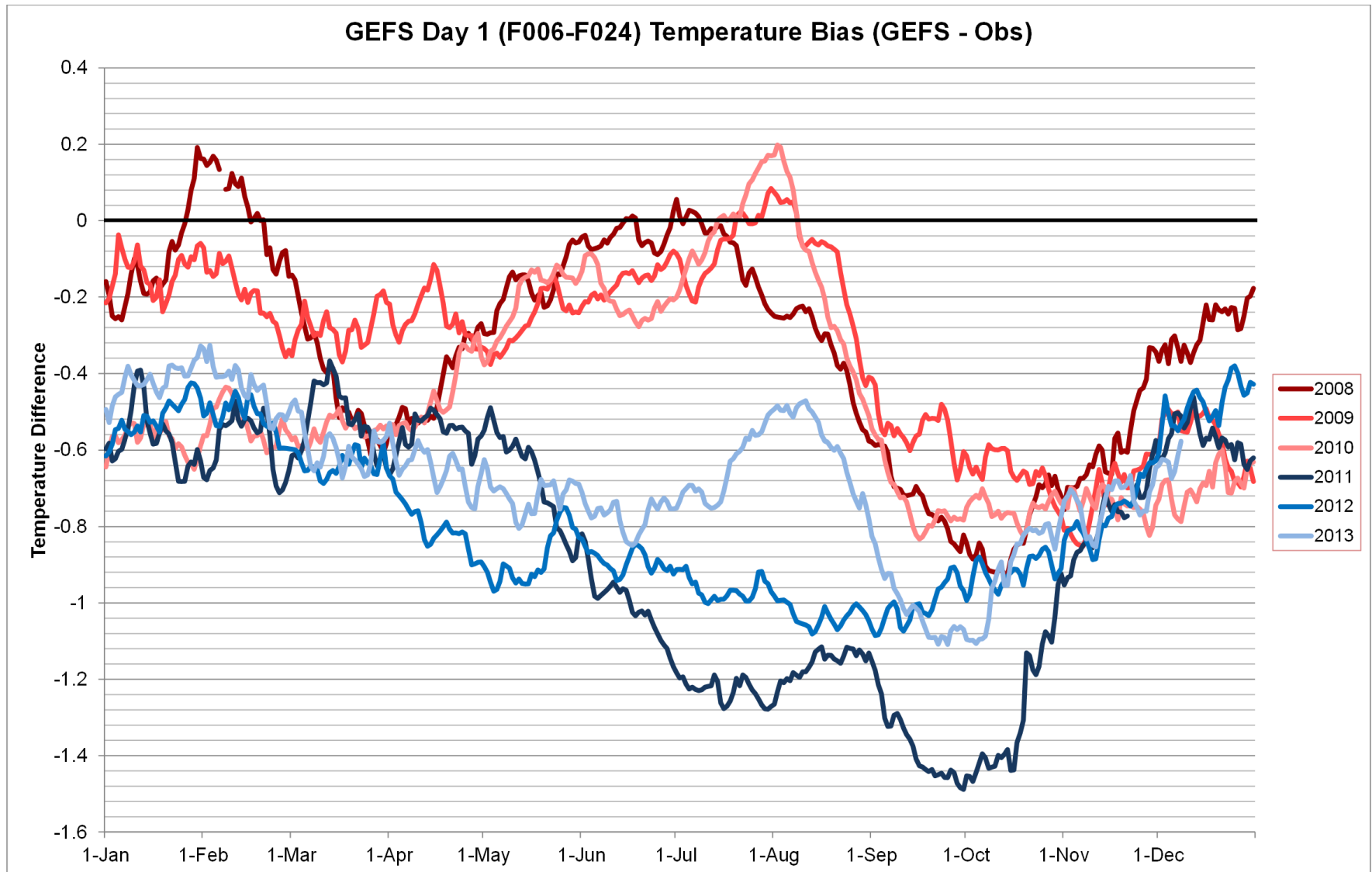
Standardized linear temperature change
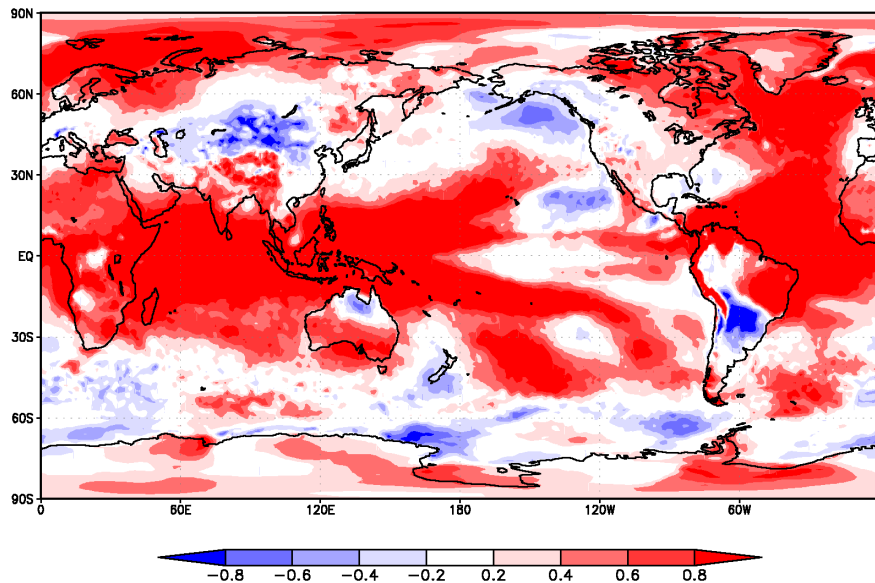
Standardized linear trend of ensemble model mean bias
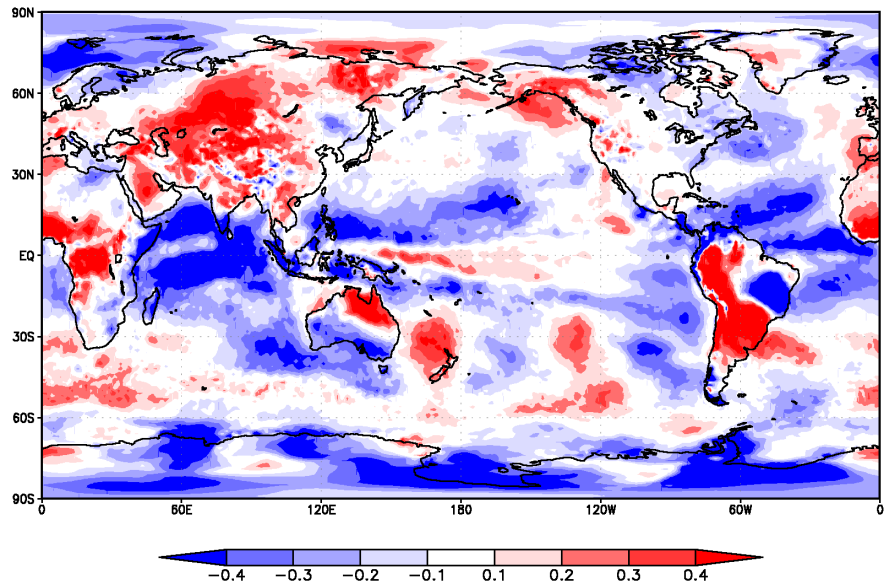
- Growing cold bias where trend is greatest?

Comparing the time series of model error from the hindcast (red) and real-time forecasts (blue) after bias-correction using hindcast period (1985-2010)

GEFS Day 1 (F006-F024) Temperature Bias (GEFS - Obs)

Attempts to correct bias in recent summer forecasts –
either trailing period (yellow) or from centered period in real-time forecast years (red) –
can have good and bad results relative to using the full hindcast alone
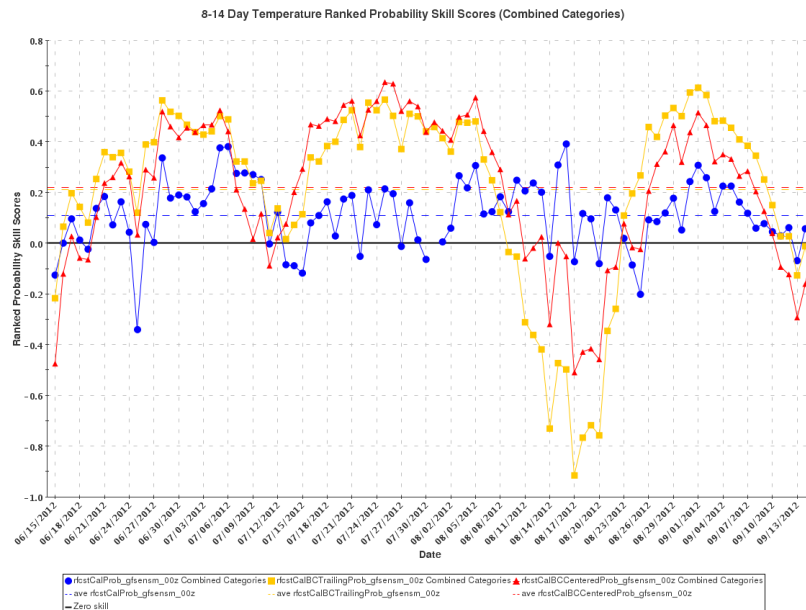
2012: Subtracting residual bias improves RPSS skill
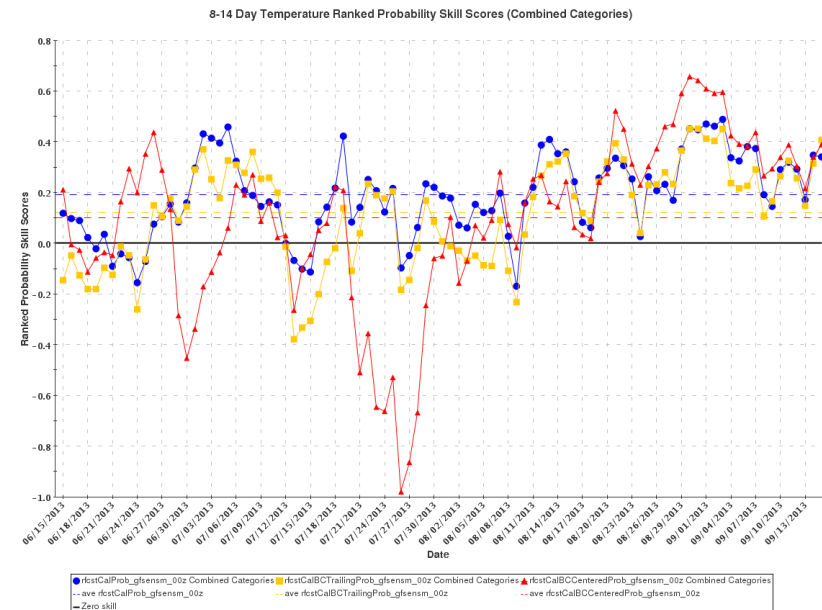
2013: Subtracting residual bias decreases RPSS skill

Summary

- Decadal temperature trend is significant fraction of subseasonal variance and of predictable signal
- The systematic bias between the ensemble mean and observations is also changing on decadal timescales.
- Regionally the predictable signal also appears non-stationary
  - This may be due to changing observation network and/or changes in climate variability
- For bias-corrections and calibration of ensemble forecasts, correction of forecasts may need to change with time
- Best practices in subseasonal may be applicable to seasonal forecasts
  - It is difficult / nearly impossible to make similar assessments of changing bias and predictability with seasonal forecasts

- Possible non-stationarity of predictability (correlation) of subseasonal variability
- Inconsistencies in hindcasts and real-time forecasts introduces additional biases;
  - However, calibration of skill (correlation) requires multi-decadal hindcasts
  - Shorter period calibrations are susceptible to variations in bias and predictability